Application of Fuzzy Multiplexing of Learning Gaussian Processes for the Interval Forecasting of Wind Speed

Miltiadis Alamaniotis [1*], Georgios Karagiannis [2]

[1] Dept. of Electrical and Computer Engineering, University of Texas at San Antonio, 1 UTSA Circle, San Antonio, TX, USA
[2] Dept. of Mathematical Sciences, Durham University, Stockton Rd, Durham, DH13LE, United Kingdom
[*]miltos.alamaniotis@utsa.edu

Abstract: Robust forecasting of wind speed values is a key element to effectively accommodate renewable generation from wind in smart power systems. However, the stochastic nature of wind and the uncertainties associated with it impose high challenge in its forecasting. In this paper, a new method for forecasting wind speed in renewable energy generation is introduced. The goal of the method is to provide a forecast in the form of an interval, that is determined by a mean value and the variance around the mean. In particular, the forecasting interval is produced according to a two-step process: in the first step, a set of individual kernel modelled Gaussian processes (GP) are utilized to provide a respective set of interval forecasts, i.e. mean and variance values, over the future values of the wind. In the second step, the individual forecasts are evaluated using a fuzzy driven multiplexer, which selects one of them. The final output of the methodology is a single interval that has been identified as the best among the GP models. The presented methodology is tested on set of real-world data and benchmarked against the individual GPs as well as the autoregressive moving average model.

Keywords: Wind speed forecasting, fuzzy logic, multiplexing, Gaussian processes, renewable power systems, machine learning.

Nomenclature

$k(x_1, x_2)$	kernel function (a.k.a. kernel)
x_x	input vector
$f(x_x)$	basis function
$f(x_x)^T$	transpose of a function
$y(x, w)$	output function
w	regression weights
$P()$	probability
$N()$	Normal distribution
σ_w^2	weight variance
I	identity matrix
C_y	covariance of function y
C	covariance matrix
$t(n)$	target for input n
ε_n	noise error
K	kernel matrix
t_N	vector of *N* targets
x_N	vector of *N* inputs
x_{N+1}	input for datapoint N+1
t_{N+1}	target for datapoint N+1
K_N	covariance matrix of the *N* observations
k	variance value of the input x_{N+1}
$\mu_A(x)$	membership function of fuzzy set A
$\{x_J, \mu_A(x_J)\}$	fuzzy singleton
$\tilde{x}$	modified input vector
$\Gamma(h_1)$	gamma function
φ, h	parameters of the kernel
R_t	real value at point t
P_t	predicted value at point t

1. Introduction

Integration of renewable sources into the power generation portfolio with prudent and alacrity, is an essential springboard for attaining sustainable energy [1]. It has been identified that one way to efficiently integrate renewables in the current power systems is via smart grid technologies. The smart grid, which couples the power infrastructure with information technologies, accommodates the intelligent management of renewable energy aiming at maximizing its utilization and minimizing its cost [2].

One of the preeminent sources of renewable energy is the wind. Wind is the driving force behind power generation from wind farms [3]. Wind farms may provide sustainable and clean energy and their integration to the power grid consists one of the grand challenges according to the US Department of Energy. The latter has set as its goal to have 20% of the total power generated by wind by 2030 [4].

One of the factors that will profoundly play a role in attaining the above goal is the utilization of data and machine learning technologies [5]. In particular, collection, storage and processing of data accommodates optimization of the decision-making process pertained to power grid operation. Furthermore, operational decisions are associated with planning and the evaluation of the cost/benefit trade-off derived from the planned actions. Regarding wind power, the parameter of interest to make optimal decisions is the wind speed at any time instance [6]. Notably, forecasting of wind speed is required to efficiently utilize the generated power.

Wind speed forecasting is performed in various time horizons that are categorized as short-, medium-, and long term [7]. Time horizons, which might differ in absolute time, are determined in order satisfy different goals pertained to grid operation. Independently of the horizon length, accuracy is of paramount importance in forecasting. However, the speed of wind depends on several unpredictable and yet unknown factors that make forecasting a highly complex and challenging problem [8].

To overcome the challenges imposed in wind speed forecasting for renewable energy generation the research community has proposed several methods. The majority of those are data driven and adopt tools from the areas of statistics and machine learning. For instance, in [9] a method that utilizes support vector machines (SVM) is discussed, while in [10] a method that combines SVM and variational mode decomposition is introduced. Neural networks have been applied and tested for wind speed forecasting as a standalone tool in [11] and [12], and in combination with wavelets in [13], and fuzzy logic in [14]. Furthermore, deep neural networks as wind speed forecasters have been proposed in [15] and [16], and the deep autoencoder in [17]. The synergism of wavelets with autoregressive integrated moving average (ARIMA) has been introduced for short term forecasting in [18], while the synergism of wavelets with SVM in [19]. Other methods that have been proposed are based on the use of polynomial autoregression [20], and Gaussian processes with particle swarm optimization (PSO) [21]. In addition, a hybrid method that uses data mining models, genetic algorithms and particle swarm optimization is described in [22], while a sophisticated method that utilizes variational mode decomposition and feedforward neural networks is presented in [23]. The aforementioned methods are only a small subset of the wind forecasting methods that exist in the literature. Admittedly, there is a plethora of forecasting methods with their own advantages and disadvantages. It should be noted that most of the proposed methods focus on providing a point forecast of the speed rather than providing an interval of values, hence overlooking the inherent uncertainty of the wind speed measurements. Except for forecasting, there are several solutions that directly handle scheduling of power generation using renewables and the imposed constraints; such a solution is given in [24], where an optimization problem is set that takes into consideration the chance constraints introduced by renewable sources. Though these solutions are adequate for providing optimal schedules, they are computationally inefficient since they require the setup of a large number of constraints that make them impractical for real time applications. Therefore, the use of forecasting is a computational efficient method for controlling and managing the power system, while handling uncertainties in a way that allows the real time management of renewable generation.

In this work, a new intelligent method that provides a forecast in the form of an interval of values over the future wind speed is proposed. A forecasting interval is defined as a range of values within it is most likely that the actual value will lie. As opposed to the point estimate, which provides one value, the interval provides more than one value. However, it is desirable in estimation problems that the interval is as narrow as possible. The interval is constructed by a multiplexer implemented as a fuzzy inference mechanism [25] that combines a set of various kernel-modelled Gaussian processes (GP) [26]. It should be noted, that the proposed method is built upon the assumption that "*the future will be similar to the most recent past*". To that end, the method implements two steps; in the first step it identifies the model that fits the observations of the most recent past and in the second step that model is used for short term forecasting [26]. The contribution of the paper contains: i) a new method for wind speed interval forecasting, ii) introduction of the novel concept of fuzzy multiplexing in wind speed forecasting, and iii) a novel framework for forecasting utilizing various individual and diverse models and thus the method contributes in building diverse ensemble models.

The roadmap of the current paper is as follows: the next section provides a background on the main components of the proposed method, while section 3 introduces the wind speed forecasting method. Section 4 provides and discusses the obtained testing results, and lastly, section 5 concludes the paper by emphasizing the main points of the paper.

2. Background

2.1. Gaussian Process Regression

One of the preeminent sets of models in machine learning is that of the kernel machines. A model is identified as a kernel machine only when it can be expressed as a function of kernel functions. A kernel function, also known as kernel, is any valid mathematical function formulated as [26]:

$$k(x_1, x_2) = f(x_1)^T f(x_2) \tag{1}$$

where $f(x)$ is a vector of basis functions. Notably, a kernel takes two values as inputs, while its form determines how the output changes with respect to the input. Therefore, it is possible to control the output of the kernel by selecting an appropriate form of kernel [26]. For instance, by assuming a linear relation of the data, then a simple choice is the linear kernel whose form is given by [27]:

$$k(x_1, x_2) = {x_1}^T x_2 \tag{2}$$

where the upper index $\cdot^T$ denotes the transpose of the input vector x_1 [26].

Gaussian process, which is a random process whose any finite subset of random variables follows a Normal multivariate distribution, may be cast into the form of a kernel machine. As kernel machines, GP may be applied to regression and classification problems. In this manuscript, GP will be utilized for regression, a scheme known as Gaussian process regression (GPR); derivation of GPR framework is briefly given below.

Below a brief introduction to the Gaussian process regression based on the Bayesian Normal linear model is given. This is because GPR can be viewed as Bayesian linear regression with a possibly infinite number of basis functions. Consider a function $y(x, w)$ expressed as a linear expansion of n bases functions $f(x) = (f_1(x), \dots, f_n(x))^T$ and unknown weights $w = (w_1, \dots, w_n)^T$ as:

$$y(x, w) = (f(x))^T w = \sum_{i=1}^{n} w_i f_i(x) \tag{3}$$

for $x \in X$.

Further, a Normal prior distribution with zero mean and covariance matrix $\Sigma_w$is assigned on the regression weights with zero mean and covariance matrix Σ_w, to account for the uncertainty of the values of the unknown weights w:

$$w \sim N(0, \Sigma_w). \tag{4}$$

Integrating out the weights in Eq. (3) with respect to the prior distribution in Eq. (4) derives the distribution of the output $y(x)$ at any input x as a Gaussian process:

$$y(\cdot) \sim GP(\mu(\cdot), k(\cdot, \cdot\,|\varphi)) \tag{5}$$

with a covariance function $k(x, x'|\varphi) = f^T(x)\Sigma_w f(x')$, and the mean function $\mu(x) = 0$, for any $x,\ x'$. Here, φ denotes the vector of the unknown parameters involved in the covariance function.

Assume that there are available N pairs of known targets and known inputs, denoted as (t_j, x_j) for $j = 1, \dots N$. In practice values $t_j = t(x_j)$ at inputs x_j are measurements, or observations which are contaminated by noise when collected. A computationally convenient assumption is to model them by considering additive Normal noise as:

$$t_j = y_j + \varepsilon_j, \quad j = 1, \dots N \tag{6}$$

where $y_j = y(x_j)$, and ε_j being the additive Normal noise with zero mean and variance equal to σ_n^2. Taking into consideration Eq. (5) and (6), the joint distribution of vector of targets $t = (t_1, \dots, t_N)$ is a Normal distribution:

$$t \sim N(0, C(\varphi, \sigma_n^2)) \tag{7}$$

where $C(\varphi, \sigma^2)$ is the $N \times N$ covariance matrix of the target values with elements:

$$[C(\varphi, \sigma^2)]_{j,j'} = k\big(x_j, x_{j'}|\varphi\big) + \sigma_n^2 \quad , j, j' = 1, \dots, N. \tag{8}$$

Unknown parameters of the covariance matrix $C(\varphi, \sigma^2)$ can be learned via maximum likelihood (ML) estimation (MLE). ML estimates $(\hat{\varphi}, \hat{\sigma}^2)$ of (φ, σ^2) are produced by maximizing the log likelihood.

$$(\hat{\varphi}, \hat{\sigma}^2) = argmax_{(\varphi, \sigma^2)} \left(-\frac{1}{2} \log\big(C(\varphi, \sigma^2)\big) \qquad -\frac{1}{2} t^T\, C(\varphi, \sigma^2) t \right) \tag{9}$$

Maximization of Eq. (9), can be performed via Newton-Raphson algorithm [28]. Here after, to easy the notation the notation $C = C(\varphi, \sigma^2)$ is adopted by suspending the dependence on the parameters.

Often interest lies in predicting the value of $t_{N+1} = t(x_{N+1})$ at a input x_{N+1}, by taking into consideration past observations. It has been shown in [26] that the predictive process is a Gaussian process:

$$y(\cdot) \sim GP(\mu^*(\cdot), k^*(\cdot,\cdot)) \tag{10}$$

with mean function:

$$\mu^*(x) = \boldsymbol{r}^{\boldsymbol{T}}(x)\, \boldsymbol{C}^{-1}\boldsymbol{t} \tag{11}$$

and covariance function:

$$k^*(x, x') = k(x, x') - r^{T}(x)\, \boldsymbol{C}^{-1}\boldsymbol{r}(\boldsymbol{x}') \tag{12}$$

where $\boldsymbol{r}^{\boldsymbol{T}}(x) = (k(x, x_1), \ldots, k(x, x_N))$, for some future input values $x,\ x'$. This implies that the predictive distribution of target $t_{N+1} = t(x_{N+1})$ at a input x_{N+1} is a Normal distribution with mean$\mu^*(x_{N+1})$ and variance $\sigma^{*,2}(x_{N+1}) = k^*(x_{N+1}, x_{N+1})$:

$$t_{N+1} | t_1, \ldots, t_N \sim N\left(\mu^*(x_{N+1}), \sigma^{*,\ 2}(x_{N+1})\right). \tag{13}$$

It is clear that both the mean and covariance functions of the predictive process/distribution Eq (11) and (12) depend on the type of the kernel function $k^*(\cdot,\cdot)$ chosen. This allows the modeler to control the output of the GPR by selecting an appropriate form of the kernel function.

An advantage of the GPR is that it allows the forming of a new kernel function able to represent complicated dependencies by combining simpler/basic kernel functions. Their combination is made by using any of the analytical operations, namely, addition, multiplication and exponential. For instance, a simple example of synthesizing a kernel by adding two simpler ones is given below:

$$k(x_1, x_2) = k_1(x_1, x_2) + k_2(x_1, x_2) \tag{14}$$

with k_1 and k_2 being two valid kernel functions [29].

2.2 Elements of Fuzzy Logic

In many practical problems, the use of classical set theory may not be enough for accurate descriptions and representation of the problem parameters. Classical set theory treats objects using a binary logic: an object either belongs or not to a set.

In contrast to the classical set theory, fuzzy logic theory assumes that the sets do not have distinct boundaries, hence, allowing a single object to partially belong to a set. In other words, an object may belong to a set up to a degree, where this degree value is evaluated by the membership function. Every fuzzy set *A* is determined by its membership function that takes the following general form [30]:

$$\mu_A(x) = [0\ 1] \tag{15}$$

where $\mu_A(x) = 0$ denotes the absence of the object *x* from the set A, while $\mu_A(x) = 1$ denotes that the object fully belongs to the set. Any other value of μ_A(x) between 0 and 1 denotes the partially participation of the object x to the set A [30].

The notion of an object being member of a set with some degree finds application to both discrete and continuous variables. In the case of discrete variables, the membership function comes in the form of pairs (called singletons in the fuzzy logic parlance [30]):

$$\mu_A(x) = \{x_1, \mu_A(x_1)\} \ldots . \{x_J, \mu_A(x_J)\} \tag{16}$$

while in the case of continuous variables the membership function takes the form of a valid analytical continuous function evaluated in the domain [0 1]. There are several shapes of functions that have been proposed for representing fuzzy sets, for instance, triangular, bell shaped, and trapezoidal. For visualization purposes, Fig. 1 shows representation of the variable x with three fuzzy sets; in that Fig. implicitly it is shown that the strength of fuzzy sets is their property of overlapping: some objects belong to more than one sets with different degrees of membership.

The representation strengths of fuzzy sets, which entail several values under the umbrella of a few functions, may also accommodate inference making. To make it clearer, representation of variables with fuzzy sets allows their correlation using a set of *IF..THEN* rules. For instance, assuming a variable *x* that is represented with fuzzy sets A, and B, and a variable y represented with sets Z, and V, then two possible rules that may be defined are the following:

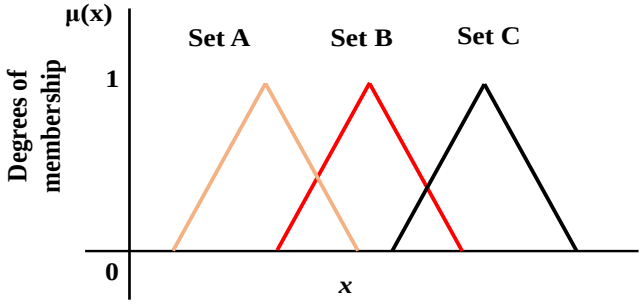

Fig. 1. *Illustrative example of fuzzy sets for the variable*

IF *x* is A, THEN *y* is V

IF *x* is B, THEN *y* is Z

where the above two rules may fully describe the relation between the two variables no matter the volume of values that each set stands for.

With the above example, it has been clearly demonstrated the representation capabilities of fuzzy logic. Adoption of fuzzy sets and the correlation of fuzzy variables via *IF..THEN* rules offers several computational benefits.

2.3 Multiplexing

A multiplexer is a term that is coined in the electronics community and has been widely used in the electric circuit domain. It is a device that has several inputs and only one input [31]. The goal of the device is to select among several inputs and forward one of them in the output. The process of selecting a single input among many is called multiplexing; a 4-to-1 multiplexing diagram is depicted in Fig. 2.

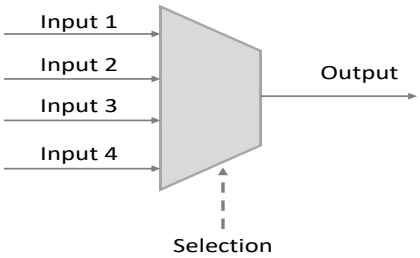

Fig. 2. *Block diagram of 4-to-1 multiplexer*

3. Wind Speed Forecasting

3.1 Problem Statement

Renewable power generation strongly relies on the inherent stochasticity of wind, and more specifically on the random factor that affect the speed of the wind. Therefore, accurate forecasting of the wind speed is a tool that will contribute in efficient management of wind sources. Furthermore, the proposed algorithm is scalable given that is it independent of the data size and the data time resolution; in other words, it may be used for any forecasting horizon.

Wind speed is the result of several natural factors that may not be fully modelled or are not fully known. Therefore, wind speed is treated as a random variable that may be forecasted by its previous values. In this manuscript, our interest lies in the very short-term forecasting and more specifically in an hour ahead of time. Performance of the forecasting method is assessed in terms of accuracy and in precision. Accuracy is measured as the distance of the predicted value to the true value, while precision is expressed in terms of forecasted variance, i.e., the width of the forecast interval.

3.2 Method

In this section, a new method for wind speed forecasting is described. A simple version of the method had been presented in [32]. The goal of the methodology is to perform very-short term forecasting that provides forecasts of wind speed in the form of intervals of bundled values.

The cornerstone of the proposed method is the multiplexing of a set of various Gaussian processes to select the GP that will provide us with the most accurate forecast. To that end, our strategy focuses on assessing each of the GP model in its most recent forecast activity and select the one that exhibited the best performance. As “best” a performance is assessed based on two factors: the accuracy of forecast, and on forecasted variance. Selection of

the GP model is performed by a fuzzy inference mechanism. The block diagram of the wind speed forecasting method is presented in Fig. 3 where all its steps are clearly shown.

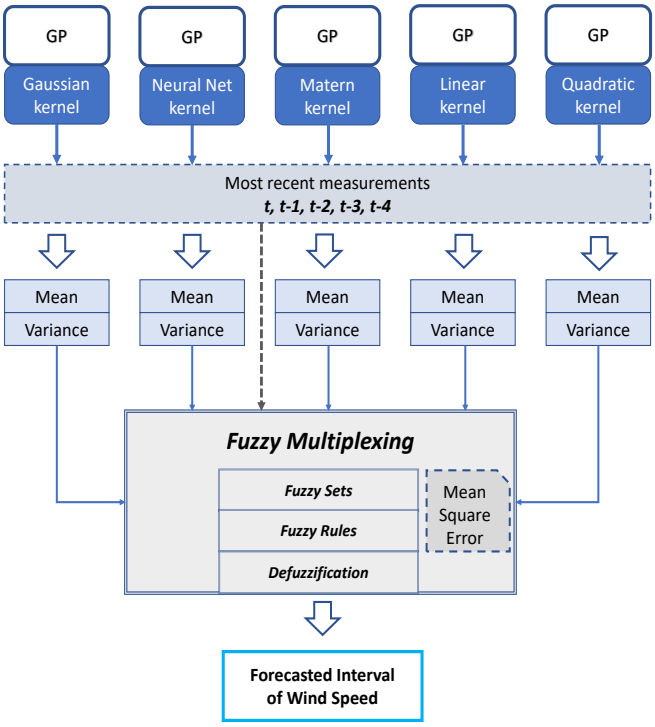

Fig. 3. *Flow chart of the proposed wind speed forecasting method for hourly ahead forecasting horizon*

Initially, a set of five Gaussian processes is adopted, where each of them is equipped with a different kernel function as depicted in Fig. 3. The analytical form of the five kernel functions utilized in the forecasting methods are given below [26]:

1. Gaussian kernel

$$k(x_1, x_2) = exp\left(-\frac{\|x_1 - x_2\|^2}{2\sigma^2}\right) \tag{17}$$

where σ^2 is a parameter evaluated using the most recent measurements.

2. Neural Net kernel

$$k(x_1, x_2) = h_0 sin^{-1}\left(\frac{2\tilde{x}_1^T \Sigma \tilde{x}_2}{\sqrt{(1 + 2\tilde{x}_1^T \Sigma \tilde{x}_1)(1 + 2\tilde{x}_2^T \Sigma \tilde{x}_2)}}\right) \tag{18}$$

where $\tilde{x} = (1, x_1, \dots, x_D)$ *i*s a modified input vector, h_0 is a scale parameter and $\boldsymbol{\Sigma}$ is the covariance matrix of the input [26].

3. Matérn kernel

$$k(x_1, x_2) = \left(\frac{2^{1-h_1}}{\Gamma(h_1)}\right)\left[\frac{\sqrt{2h_1}|x_1 - x_2|}{h_2}\right]^{h_1} \times K_{h_1}\left(\frac{\sqrt{2h_1}|x_1 - x_2|}{h_2}\right) \tag{19}$$

modelled as a function of two positive hyperparameters denoted as h_1, h_2, and the modified Bessel function K_{h_1}.

4. Linear kernel

$$k(x_1, x_2) = h\, {x_1}^T x_2 \tag{20}$$

comprised of a scale hyperparameter denoted as h.

5. Quadratic kernel

$$k(x_1, x_2) = (x_1{}^T x_2)^2 \tag{21}$$

which does not include any parameters.

As shown in Fig. 3, in the next step, the five GP models are exposed to the most recent 5 wind speed measurements. Selection of the most 5 measurements is essentially based on previous work as presented in [21], [32] and [33]. Those measurements are utilized for evaluating the parameters of the kernel, if needed, a process known as training. Furthermore, the trained GP are utilized to reproduce the learned measurements. The reproduced measurements are provided in the form of a mean and a variance for each of the five measurements.

According to Fig. 3, the obtained mean and variance values are forwarded to the *fuzzy multiplexing module*. The multiplexer collects those values and assesses the degree of reproduction: the highest the degree is, the more accurate is the learning. Therefore, the multiplexer selects the GP model that provides the highest degree of reproduction. In other words, the selection of GP made by the multiplexer is based on the degree of reproduction.

As shown in the lower part of Fig. 3, the multiplexing selection module is implemented as a fuzzy inference engine. The first part of the fuzzy inference encompasses fuzzy modelling of two variables via adoption of appropriate fuzzy sets. The variables utilized for selection making are, namely, the accuracy and the precision variables.

Regarding accuracy, the mean average percentage error (MAPE) measure is adopted (see also Fig. 3) to quantify the distance between the five most recent measurements and their reproduced means as taken by the GP model. In particular, the MAPE takes the following form:

$$MAPE = \frac{100}{M}\sum_{t=1}^{M}\left|\frac{R_t - P_t}{R_t}\right| \tag{22}$$

with *M* being equal to 5 that is equal to the population of the most recent measurements. The range of values of MAPE lies in the interval $[0\ 100]$ (which is a percentage range), which is spanned by three fuzzy sets as shown in Fig. 4.

Furthermore, a group of three fuzzy sets is also selected for representing precision. In this work, precision is computed as the mean variance of the reproduced variance values for the most recent measurements as obtained by the GP models:

$$Precision = \frac{1}{M}\sum_{t=1}^{M} V_t^G \tag{23}$$

where V_t^G denotes the computed variance of the model *G* for the measurement t.

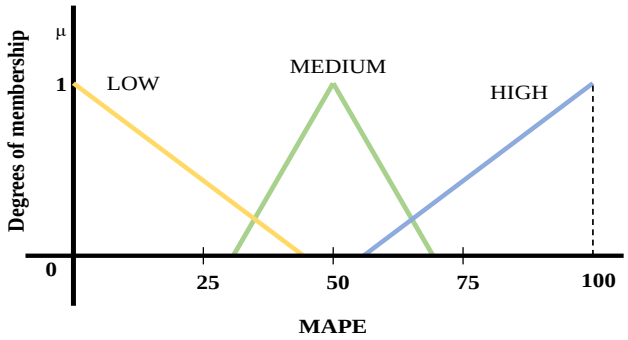

Fig. 4. *Fuzzy set representation of the variable MAPE that represents accuracy*

The fuzzy representation of the variable precision is depicted in Fig. 5. Notably, the range of values of the variable precision is $[0, \infty]$ since precision is a positive measure.

In addition to the above variables, it is also defined the fuzzy variable *reproduction*. This variable is essential in implementing the fuzzy inference engine given that it plays the role of the output variable of the system as shown in Fig. 3. Its range of values coincides with the interval [0 1], and denotes the degree of successful reproduction of the measurements by the GP model. In our approach, the higher the value is, the more successful the reproduction is. Similarly to the other two variables, the variable *reproduction* is also fuzzified using three fuzzy sets (LOW, MEDIUM, HIHG); its fuzzification is depicted in Fig. 6.

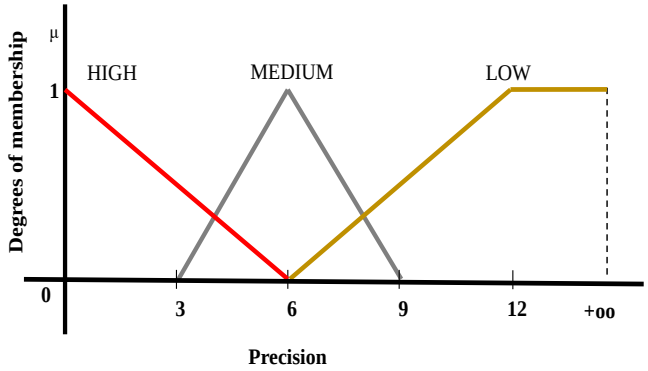

Fig. 5. *Fuzzy set representation of the variable precision*

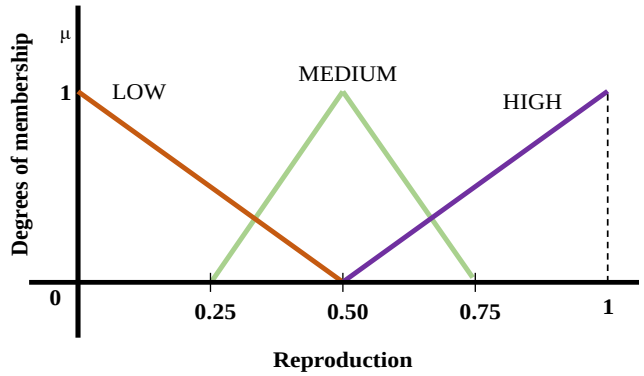

Fig. 6. *Fuzzy set representation of the output variable reproduction*

Having defined the three variables, i.e. two inputs and one output, then we proceed by developing the body of the fuzzy inference, i.e. the fuzzy rules (as shown in the lower part of Fig. 3). The goal of the fuzzy inference is to correlate the fuzzy sets of the inputs with the sets of the output variable [29]. The fuzzy rules that comprise the inference engine in the current work are given below:

1) *IF* MAPE is LOW, and Precision is HIGH,
THEN Reproduction is HIGH

2) *IF* MAPE is LOW, and Precision is MEDIUM,
THEN Reproduction is MEDIUM

3) *IF* MAPE is LOW, and Precision is LOW,
THEN Reproduction is MEDIUM

4) *IF* MAPE is MEDIUM, and Precision is HIGH,
THEN Reproduction is LOW

5) *IF* MAPE is MEDIUM, and Precision is MEDIUM,
THEN Reproduction is MEDIUM

6) *IF* MAPE is MEDIUM, and Precision is LOW,
THEN Reproduction is MEDIUM

7) *IF* MAPE is HIGH, and Precision is HIGH,
THEN Reproduction is MEDIUM

8) *IF* MAPE is HIGH, and Precision is MEDIUM,
THEN Reproduction is LOW

9) *IF* MAPE is HIGH, and Precision is LOW,
THEN Reproduction is LOW

where the population of the rules is equal to 9. Notably, the fuzzy rules have two conditions in the left-hand side, i.e. one for each input variable, while there is only one statement in the right-hand side (because there is a single output). It should be noted that the rules are developed based upon the linguistic terms of the fuzzy sets, hence, expressing

the semantic correlations among the variables. The semantic correlation and subsequently the development of rules is conducted empirically based on intuition and the modeler's strategy – for instance a conservative strategy would require the correlation of a few fuzzy sets for every variable while correlate "medium" sets with the "low" ones-. Though the rules are written in linguistic terms, they are numerically evaluated. In the current work, the rules are evaluated using the *Mamdani Min* implication operator [30] whose analytical form is given by:

$$\varphi_c\big(\mu_A(x), \mu_B(y)\big) = min\big(\mu_A(x), \mu_B(y)\big) \quad (24)$$

where A and B are fuzzy sets, and x, y are the input and output values respectively. In this work, the fuzzy rules have two conditions in the left-hand right side connected with the operator AND. With respect to fuzzy logic, the operator AND is numerically evaluated as the maximum of the two membership values:

$$A(x)\ AND\ B(x) = max\big(\mu_A(x), \mu_B(x)\big) \quad (25)$$

where *A*(*x*) and *B*(*x*) are two fuzzy sets (or fuzzy statements, see [29]).

In a fuzzy inference engine, it is common that more than one rule is fired, i.e. multiple rules are valid at the same time, and as a result the output of the engine is a fuzzy set. However, in practical problems, a single value is required instead of a fuzzy output. Hence, a defuzzification step is adopted to transform the fuzzy output into a single value, known as crisp value [30]. In this work, the *Center of Area defuzzification* method is adopted whose analytical form is given by [30]:

$$y = \frac{\sum_{i=1}^{N} y_i \mu_{out}(y_i)}{\sum_{i=1}^{N} \mu_{out}(y_i)} \quad (26)$$

with y_i being the elements of the fuzzy set and N denotes the element population. The value *y* computed by Eq. (26) is the final output of the fuzzy inference mechanism. The fuzzy inference is iterated five times - equal to population of GP models – providing a set of five values.

The set of five values, which represent the degree of reproduction achieved by the individual GP models, are ordered in an ascending order. The model with the highest degree is selected as the final GP model that will be utilized for hour ahead wind speed forecasting. It should be noted, that in case of two or more models provide the same degree of reproduction, then the model with lowest MAPE is selected.

In sum, the fuzzy multiplexer selects one out of five models, to subsequently use it to obtain the next hour forecast. The presented method is applied every one hour aiming at selecting the most suitable forecasting model at each time. Overall, the proposed method aspires to capture the most recent wind speed dynamics by assessing the most recent observed data [32].

4. Application to Wind Data

4.1. Testing Setup

The presented wind speed forecasting method is tested on a set of real-world datasets taken from the National Renewable Energy Laboratory (NREL) located in Colorado, USA [34]. In the current work, the hourly wind speed data measured by NREL for day times 4am-8pm in the annual period of *January 1, 2016 - December 31, 2016* is used. The following process was followed for obtaining the results: the training data was comprised of the most five hourly wind speed observed values, before the targeted hour as mentioned in the previous section.

The performance of the method is assessed on two dimensions: i) the accuracy of prediction quantified as the error between the interval mean and the actual value, and ii) the precision expressed as the average interval width achieved by the method. The proposed method is benchmarked against the individual GP models as well as a statistical model and more particularly the autoregressive moving average (ARMA) model.

The results are presented on the average per month accuracy and interval width. The accuracy is computed as the Mean Square Error (MSE) given by:

$$Accuracy = \frac{1}{\#D} \sum_{d=1}^{\#D} (P_d - R_d)^2 \quad (27)$$

with P_d and R_d being the forecasted and real value respectively, while *#D* denotes the number of days per month. Likewise, the precision is taken as the interval width is taken as the average per month interval width. Lastly, it should be noted that the ARMA forecasted was taken as ARMA(4,1) with the coefficients being evaluated with the Akaike Information Criterion (AIC).

4.2 Results

In this section, the presented method is applied on the wind speed data and the obtained results are obtained. Initially, the average per month MSE obtained for the testing period January 2016 – December 2016 are presented in Fig. 7-18.

Obtained MSE results demonstrate that there is no dominant forecaster, i.e. there is not a single forecaster that consistently provides the lowest error. In the first semester, the ARMA forecaster marginally provides the lower error for four months, However, for the second semester, ARMA performance significantly deteriorates by providing high error forecasts for all six months.

With regard to the individual GP forecasters, a general observation shows that all of them provide low MSE values, while their error values are very close (i.e. within 2 units). In particular, the GP-Quadratic is the best performer for month August, and GP-linear for months July and December.

For the rest six months – February, June, August, September, October, November- the lowest error is obtained by the proposed GP-Fuzzy method. It should be noted, that even in the cases that our method is not the best performer, it still provides low error values that are close to that respective best performer.

By focusing only on the GP forecasters (leaving out ARMA) – the individual ones as well the GP-Fuzzy method – it is observed that none of them prevails over the others with respect to MSE. Furthermore, the forecasters seem to have variations in their performance; for example, the GP-NN provides low errors for 5 months and very high for the rest of them. Similar variation is observed for the rest of them with the most characteristic example to the be the GP-linear.

As opposed to the single GP forecasters, the proposed GP-Fuzzy presents a more stable performance. In particular, it is observed that the GP-Fuzzy consistently provides low error and its performance is within the lowest half at all tested cases. Notably, when all the individual GP forecasters provide performance that is close to each other, the GP-fuzzy method provides the best accuracy (i.e., the lowest error). This observation emerges the implicit dependency of the GP-fuzzy method with the five individual GP models. Therefore, our method provides a more stable forecasting performance over a long period of time (12 months - overall 6300 hourly forecasts), a property that is desirable in any type of application. Notably, the proposed GP-Fuzzy is a robust method that provides low error, and as a result, it eliminates the requirement to blindly select (or perform extensive and intensive trial and error testing for selecting) a specific kernel form for our kernel machine forecaster; selection is difficult given that their forecasting performance is not known a priori. Furthermore, the presented GP-fuzzy method handles the forecasting diversity: every forecaster may be suitable for a specific dataset or for a specific set of data properties. Hence, the advantage of the proposed method is that it provides a method for encompassing the ability of forecaster selection based on some recent evidence (i.e., recent measurements) in a fast, transparent and automated way.

With regard to forecasted interval, the results are presented in Table 1 and Table 2. In particular, the average per month width of forecasted interval is presented as it is computed by the GP models. It should be noted that the ARMA model is not included in Tables 1-2 given that it provides a point estimation and not an interval of forecasts. This may be a considered as shortcoming of ARMA because interval forecasting allows the system operator to account for last minute unexpected changes in the operation of the system. Observation of Tables 1-2 shows that, with the exception of GP equipped with a Neural Net kernel, the rest forecasters provide intervals with widths of similar values. Though the differences in their width are small, it should be noted that there is not a single dominant GP forecaster that consistently provides the narrowest interval. Notably, our method provides widths in the narrow end of the forecasters list. Similar to MSE criterion, the GP-fuzzy method also exhibits a stable performance with respect to interval widths. At this point, the strength of the proposed method once more is emphasized: given that it is not known a priori which of the models will provide the best performance, hence, the best

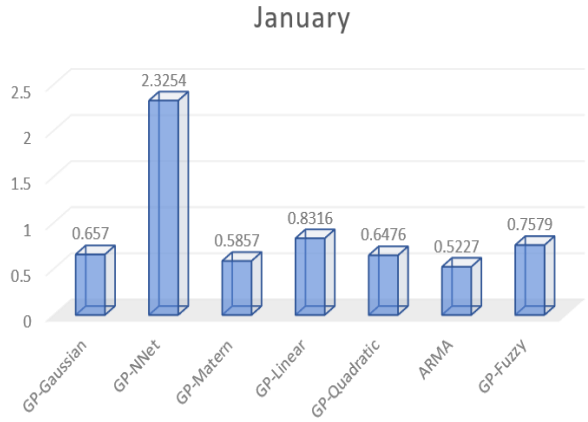

Fig. 7. *Monthly average MSE for January 2016*

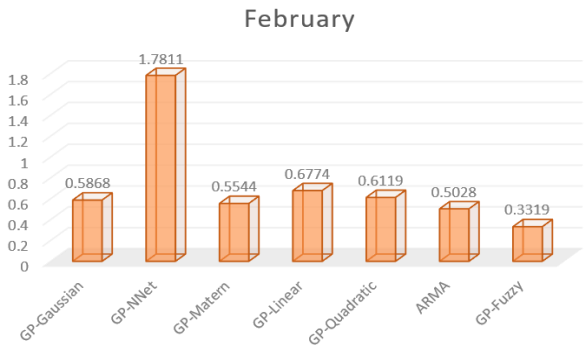

Fig. 8. *Monthly average MSE for February 2016*

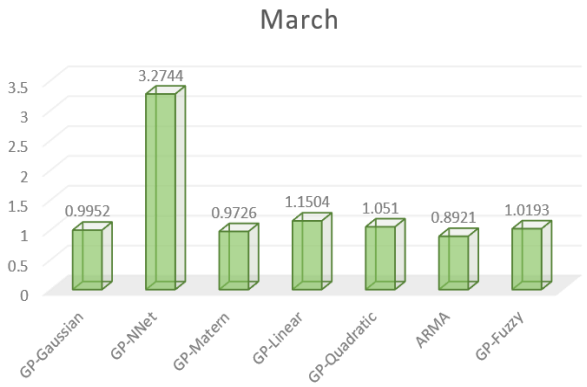

Fig. 9. *Monthly average MSE for March 2016*

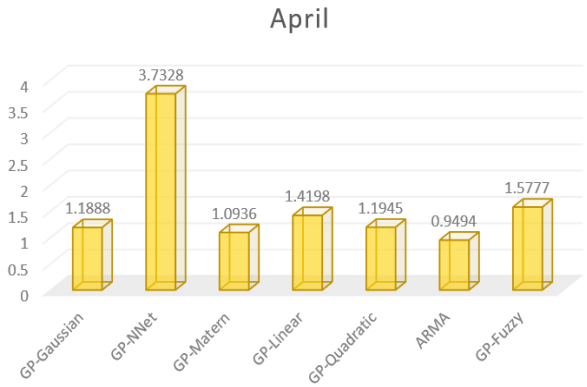

Fig. 10. *Monthly average MSE for April 2016*

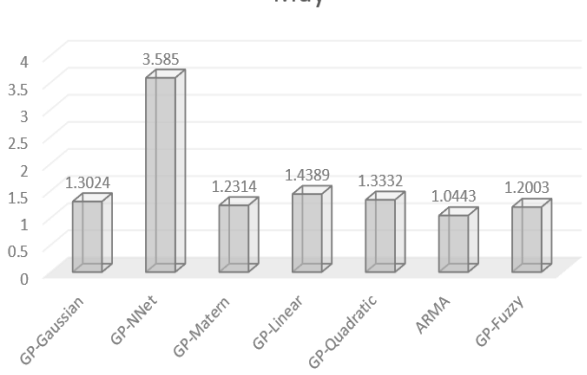

Fig. 11. *Monthly average MSE for May 2016*

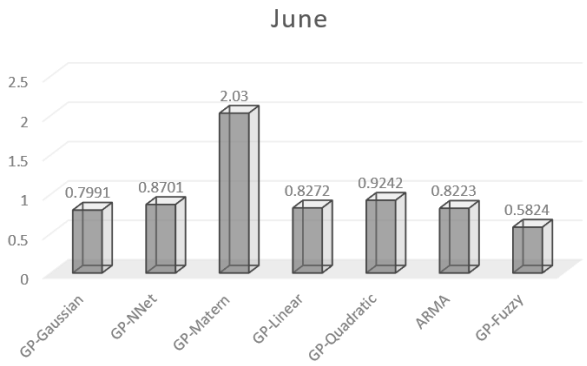

Fig. 12. *Monthly average MSE for June 2016*

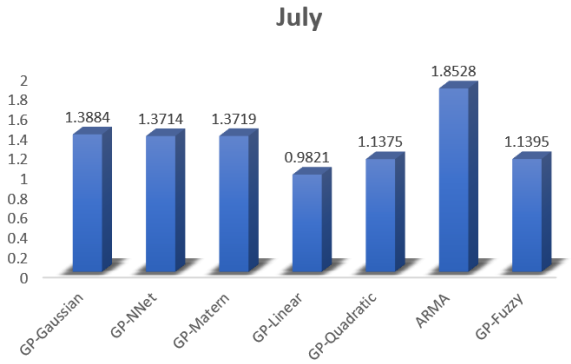

Fig. 13. *Monthly average MSE for July 2016*

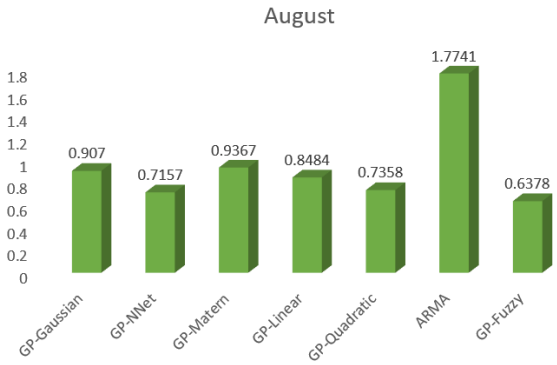

Fig. 14. *Monthly average MSE for August 2016*

Table 1 Average per month interval width (m/h) for the GP forecasters (January-June 2016)

	Jan	Feb	Mar	Apr	May	Jun
GP-Gaussian	2.84	2.43	2.97	3.25	3.29	2.82
GP-NNet	6.79	5.63	7.29	7.98	7.86	5.87
GP-Matérn	3.06	2.58	3.21	3.49	3.55	3.01
GP-Linear	3.59	2.99	3.71	4.17	4.17	3.54
GP-Quadratic	2.81	2.41	2.94	3.22	3.25	2.77
GP-Fuzzy	3.26	2.61	3.40	4.01	3.76	3.10

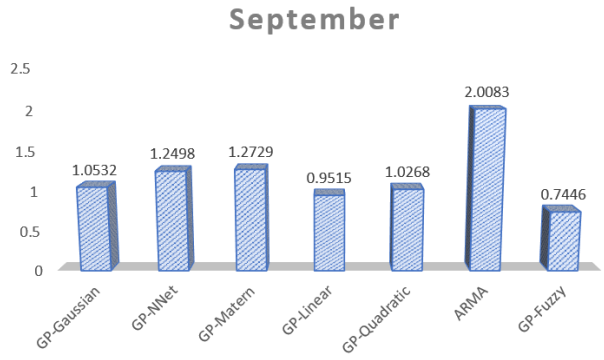

Fig. 15. *Monthly average MSE for September 2016*

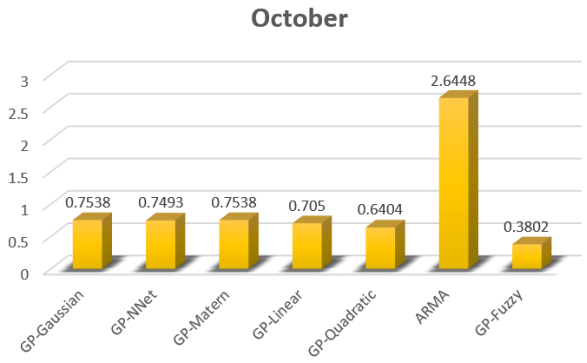

Fig. 16. *Monthly average MSE for October 2016*

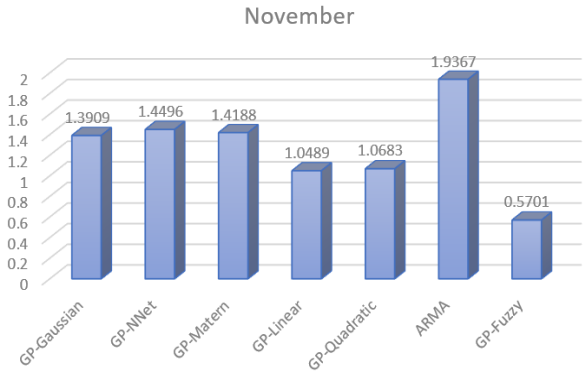

Fig. 17. *Monthly average MSE for November 2016*

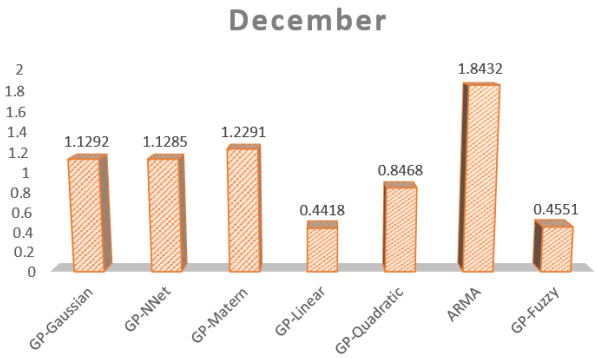

Fig. 18. *Monthly average MSE for December 2016*

model cannot be ahead of time selected. However, adoption of the proposed GP-fuzzy method secures a robust forecasting output with respect to error and the width of the forecasted interval, while avoiding the risk of selecting the forecaster.

Table 2 Average per month interval width (m/h) for the GP forecasters (July-December 2016)

	Jul	**Aug**	**Sep**	**Oct**	**Nov**	**Dec**
GP-Gaussian	6.43	5.29	6.11	4.75	6.78	5.63
GP-NNet	6.37	4.22	6.10	4.74	6.78	5.63
GP-Matérn	6.43	5.29	6.11	4.75	6.78	5.63
GP-Linear	4.14	3.59	3.72	3.08	3.54	3.35
GP-Quadratic	5.31	4.36	5.08	3.87	5.26	4.56
GP-Fuzzy	2.39	3.74	2.74	2.73	3.24	3.11

To further evaluate the forecasters, it is also computed the percentage of actual points that fall within the forecasted intervals. The results in the form of percentages are presented in Table 3 and Table 4 (marked as *encompass percentage*) for the whole 12-month long test period. It is clear that the presented methodology outperforms the other forecasters; in particular it encompasses 14% more values than the best of the rest forecasters in the first semester and 1% for the second semester of 2016; hence for the whole year the difference in the encompass ratio is about 7.5%.

To better assess the ability of forecasters to encompass the actual values with respect to the forecasted width and emphasize the capabilities of the proposed method, we also compute the following ratio for each of the GP forecasters:

$$Encompass\ Rate = \frac{\#E}{W_F^A} \tag{28}$$

where $\#E$ is the population of actual wind speed values that lie within the forecasted intervals, and W_F^A is the average width of the interval for forecaster *F*. The higher the value of the above ratio, the better the performance of the method. The goal of the encompass rate is to show that the more efficient interval forecaster includes the highest number of actual values with the narrowest width. This can be understood by considering that among two forecasters whose intervals contain the same number of actual values (i.e., #E in Eq. (28)), the most efficient is the one that provides narrower intervals (i.e., the denominator in Eq. (28)). Thus, for forecasters with the same *E,* the encompass ratio is higher when the denominator is smaller: In fact, this shows that the same amount of actual values were correctly forecasted by smaller intervals (less uncertainty) which is our desired goal. The values obtained for each forecaster by Eq. (28) are also provided in Tables 3-4, where it is clearly demonstrated that the GP-fuzzy method provides the higher *Encompass Ratio* among all the tested methods. The above result, clearly highlights the ability of our method to provide high accuracy with narrow intervals.

Table 3 Performance metrics of forecasters with respect to actual values encompassed in the forecasted intervals for period *January 2016 – June 2016*

Forecaster	Encompass Percentage	Encompass Ratio
GP-Gaussian	48.2%	16.40
GP-NNet	40%	5.79
GP-Matérn	51.38%	16.28
GP-Linear	47.88%	12.93
GP-Quadratic	47.98%	16.51
GP-Fuzzy	64.9%	19.32

Table 4 Performance metrics of forecasters with respect to actual values encompassed in the forecasted intervals for period *July 2016 – December 2016*

Forecaster	Encompass Percentage	Encompass Ratio
GP-Gaussian	42.90%	7.36
GP-NNet	42.70%	7.58
GP-Matérn	42.81%	7.44
GP-Linear	45.61%	12.80
GP-Quadratic	37.20%	7.84
GP-Fuzzy	46.06%	15.40

For visualization purposes, the forecasted intervals of the first week of February (Feb 1-7) are depicted in Fig. 19. In addition, the real wind speed values in the same Fig. are plotted, where the vast majority of actual values lie within the forecasted wind speed intervals. Furthermore, the forecasted intervals of the last week of April are depicted in Fig. 20 and the first week of June in Fig. 21 respectively. Notably, it is also observed that the GP-fuzzy intervals encompass the majority of the actual values. Fig. 19-21 demonstrate the forecasting ability, and more specifically highlight the ability of the with the GP-fuzzy method to provide narrow width intervals. The main conclusion by inspecting Figs. 19-21 is that our method is able to capture the trend of the actual data. In particular, Figs. 19-21 confirm the ability of the GP-fuzzy forecaster to provide narrow intervals, while being able to capture the dynamics of the wind speed (i.e., following the data trend). It should be noted that the data not encompassed in the forecasted intervals seem to be outliers as compared to the rest nearby in time values; the latter is apparent in Fig. 21 where the actual value at time around 100 is far beyond the bulk volume of values around that time - and thus the GP-fuzzy fails to encompass it-. Finally, we observe that the presented method is able to capture the abrupt transitions in the trend of wind speed values as it is apparent in Fig. 20.

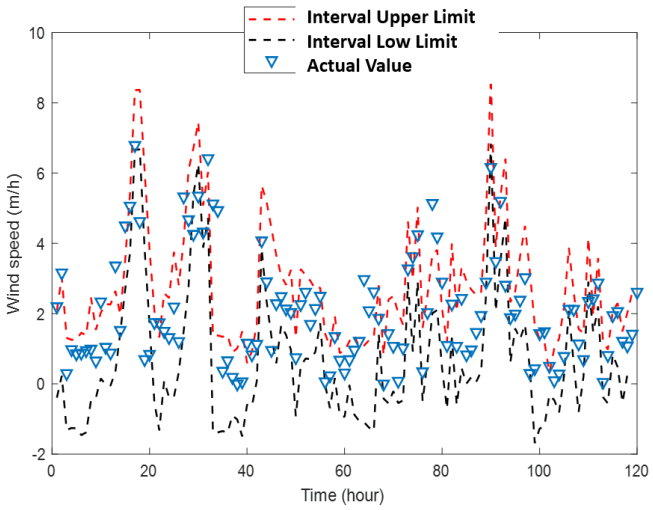

Fig. 19. *Plot of forecasted interval vs. real wind speed values of testing period February 1-7, 2016*

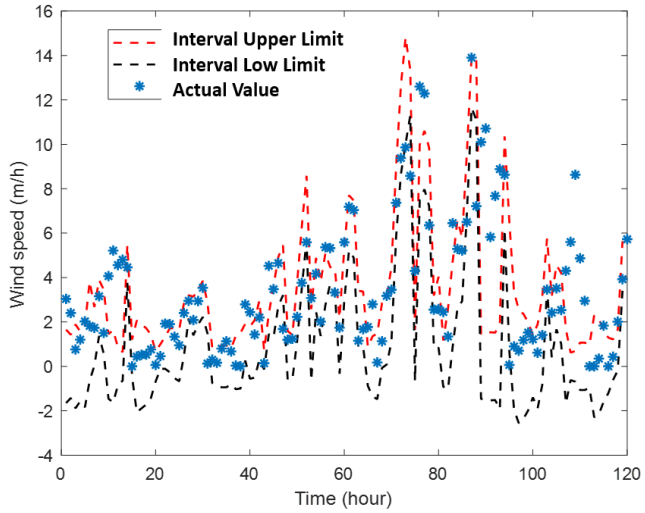

Fig. 20. *Plot of forecasted interval vs. real wind speed values of testing period April 24-30, 2016*

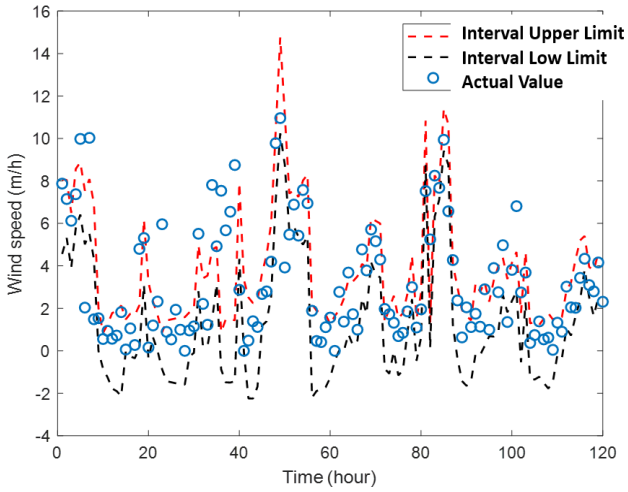

Fig. 21. *Plot of forecasted interval vs. real wind speed values of testing period June 1-7, 2016*

In sum, the presented GP-Fuzzy wind speed methodology provides a forecast in the form of an interval of values. The method is able to provide a robust forecasting as shown from the obtained results. Our method not only provides accurate forecasting, but it also provides narrow intervals – in other words, a forecast with low uncertainty – of hourly ahead of time wind speed values.

5. Conclusion

In this paper, a new robust method for wind speed forecasting has been proposed. The cornerstone of the proposed forecaster is the adoption of a set of Gaussian processes equipped with various kernel functions, and their subsequent multiplexing using a fuzzy inference system. The fuzzy multiplexing system is comprised of a set of 9 fuzzy rules and its goal is to use the most recent observed wind speed values and provide a value – known as degree of reproducibility- for each of the individual GP. These values are used to select the GP model that will be likely the most accurate forecaster of the next hour wind speed value.

The proposed method has been tested on a set of real-world wind speed measurements. The testing period consisted of the hourly values for a period of six months providing a set of over 6300 forecasted values. The GP-fuzzy method was compared to the five single GP models as well as to ARMA model. Performance was tested with respect to accuracy and width of forecasted interval. Our forecaster, though it did not provide the best performance for all tested cases, exhibited a robust behaviour by consistently providing low error and narrow intervals. Furthermore, it clearly outperformed all the rest forecasters by being the model that encompasses the highest number of actual values in the interval, while keeping the interval width narrow. The latter performance was demonstrated by the *Encompass Ratio* that was the highest among the tested forecasters.

6. References

[1] Camioto, F. D. C., Mariano, E. B., Santana, N. B., Yamashita, B. D., Rebelatto, D. A. D. N.: 'Renewable and sustainable energy efficiency: An analysis of Latin American countries', Environmental Progress & Sustainable Energy, 2018, 37, (6), pp. 2116-2123

[2] Kupzog, F., Sauter, T., Pollhammer, K.: 'IT-enabled integration of renewables: A concept for the smart power grid,' EURASIP Journal on Embedded Systems, 2011, 2011, (1), pp. 737543 (1-8)

[3] Ladenburg, J., Lutzeyer, S.: 'The economics of visual disamenity reductions of offshore wind farms—Review and suggestions from an emerging field,' Renewable and Sustainable Energy Reviews, 2012, 16, (9), pp. 6793-6802

[4] NREL, '20% by Wind Energy by 2030' (US D.O.E., 2009), pp. 1-248

[5] Marvuglia, A., Messineo, A.: 'Monitoring of wind farms' power curves using machine learning techniques', Applied Energy, 2012, 98, pp. 574-583

[6] Wang, X., Jiang, C., Liu, Y., Wang, J.: 'Decision-making Model and Method for Spinning Reserve and Risk of Power Systems Integrated with Largescale Wind Farms', Automation of Electric Power Systems, 2014, 38, (13), pp. 64-70

[7] Soman, S.S., Zareipour, H., Malik, O., Mandal, P.: 'A review of wind power and wind speed forecasting methods with different time horizons', Proc. IEEE North American Power Symposium, Arlington, TX, USA, September 2010, pp. 1-8

[8] De Giorgi, M.G., Ficarella, A., Tarantino, M.: 'Error analysis of short-term wind power prediction models', Applied Energy, 2011, 88, (4), pp. 1298-1311

[9] Mohandes, M.A., Halawani, T.O., Rehman, S., Hussain, A.A.: 'Support vector machines for wind speed prediction', Renewable Energy, 2004, 29, (6), pp. 939-947

[10] Wang, X., Yu, Q., Yang, Y.: 'Short-term wind speed forecasting using variational mode decomposition and support vector regression', Journal of Intelligent & Fuzzy Systems, 2018, 34, (6), pp. 3811-3820

[11] Cardenas-Barrera, J.L., Meng, J., Castillo-Guerra, E., Chang, L.: 'A neural network approach to multi-step-ahead, short-term wind speed forecasting', Proc. 12th IEEE International Conference on Machine Learning and Applications, Miami, FL, USA, December 2013, vol. 2, pp. 243-248

[12] Huang, S.H., Mu, K.M., Lu, P.Y., Tsao, C.Y., Leu, Y. G., Chou, L.F.: 'The application of neural network in wind speed forecasting', Proc. IEEE 12th International Conference on Networking, Sensing and Control, Taipei, Taiwan, April 2015, pp. 366-370

[13] Meng, A., Ge, J., Yin, H., Chen, S.: 'Wind speed forecasting based on wavelet packet decomposition and artificial neural networks trained by crisscross optimization algorithm', Energy Conversion and Management, 2016, 114, pp. 75-88

[14] Ma, X., Jin, Y., Dong, Q.: 'A generalized dynamic fuzzy neural network based on singular spectrum analysis optimized by brain storm optimization for short-term wind speed forecasting', Applied Soft Computing, 2017, 54, pp. 296-312

[15] Khodayar, M., Wang, J., Manthouri, M.: 'Interval deep generative neural network for wind speed forecasting', IEEE Transactions on Smart Grid, 2018, 10, (4), pp. 3974-3989

[16] Khodayar, M., & Wang, J.: 'Spatio-temporal graph deep neural network for short-term wind speed forecasting', IEEE Transactions on Sustainable Energy, 2018, 10, (2), pp. 670-681

[17] Mezaache, H., Bouzgou, H.: 'Auto-Encoder with Neural Networks for Wind Speed Forecasting', Proc. IEEE International Conference on Communications and Electrical Engineering (ICCEE), El Oued, Algeria, December 2018, pp. 1-5

[18] Singh, S. N., Mohapatra, A.: 'Repeated wavelet transform based ARIMA model for very short-term wind speed forecasting', Renewable Energy, 2019, 136, pp. 758-768

[19] Ni-ya, C., Zheng, Q., Xiao-feng, M.: 'A hybrid model for short-term wind speed forecasting based on wavelet and Support Vector Machine', Proc. IET Conference on Renewable Power Generation, Edinburgh, UK, September 2011, pp. P24

[20] Karakuş, O., Kuruoğlu, E.E., Altınkaya, M. A.: 'One-day ahead wind speed/power prediction based on polynomial autoregressive model', IET Renewable Power Generation, 2017, 11, (11), pp. 1430-1439

[21] Alamaniotis, M., Karagiannis, G.: 'Integration of Gaussian Processes and Particle Swarm Optimization for Very-Short Term Wind Speed Forecasting in Smart Power', International Journal of Monitoring and Surveillance Technologies Research, 2017, 5, (3), pp. 1-14

[22] Wang, J., Zhang, F., Liu, F., Ma, J.: 'Hybrid forecasting model-based data mining and genetic algorithm-adaptive particle swarm optimisation: a case study of wind speed time series', IET Renewable Power Generation, 2016, 10, (3), pp. 287-298

[23] Gendeel, M., Yuxian, Z., Aoqi, H.: 'Performance comparison of ANNs model with VMD for short-term wind speed forecasting', IET Renewable Power Generation, 2018, 12, (12), pp. 1424-1430

[24] Li Y., Yang Z., Li G., Zhao D., Tian W.: 'Optimal scheduling of an isolated microgrid with battery storage considering load and renewable generation uncertainties', IEEE Transactions on Industrial Electronics, 2018, 66, (2), pp. 1565-1575

[25] Alamaniotis, M., Agarwal, V.: 'Fuzzy Integration of Support Vector Regression Models for Anticipatory Control of Complex Energy Systems', International Journal of Monitoring and Surveillance Technologies Research, 2014, 2, (2), pp. 26-40

[26] Rasmussen, C. E., Williams, C.: 'Gaussian Processes for Machine Learning' (MIT Press, 2006)

[27] Alamaniotis, M., Agarwal, V., Jevremovic, T.: 'Anticipatory monitoring and control of complex energy systems using a fuzzy based fusion of support vector regressors', Proc. 5th IEEE International Conference on Information, Intelligence, Systems and Applications, Chania, Greece, July 2014, pp. 33-37

[28] Süli, E., Mayers, D. F.: 'An introduction to numerical analysis' (Cambridge university press, 2003)

[29] Alamaniotis, M., Tsoukalas, L.H.: 'Multi-kernel anticipatory approach to intelligent control with application to load management of electrical appliances', Proc. 24th IEEE Mediterranean Conference on Control and Automation (MED), Athens, Greece, June 2016, pp. 1290-1295

[30] Tsoukalas, L.H., Uhrig, R.E.: 'Fuzzy and Neural Approaches in Engineering' (Wiley and Sons Ltd., 1997)

[31] Jiang, Y., Al-Sheraidah, A., Wang, Y., Sha, E., Chung, J.G.: 'A novel multiplexer-based low-power full adder', IEEE Transactions on Circuits and Systems II: Express Briefs, 2004, 51, (7), pp. 345-348

[32] Alamaniotis, M., Karagiannis, G.: 'Learning Uncertainty of Wind Speed Forecasting using a Fuzzy Multiplexer of Gaussian Processes', Mediterranean Conference on Power Generation, Transmission, Distribution, and Energy Conversion (MEDPOWER 2018), Dubrovnik, Croatia, November 12-15, 2018, pp. 1-6

[33] Alamaniotis, M., Karagiannis, 'Minute Ahead Wind Speed Forecasting Using a Gaussian Process and Fuzzy Driven Assimilation', Proc. *IEEE Milan PowerTech*, Milano, Italy, June 23-27, 2019, pp. 1-6.

[34] Andreas, A., Wilcox, S.: 'Observed atmospheric and solar information system (oasis); Tucson, Arizona (data)' NREL Report No. DA-5500-56494, 2010